# structural bioinformatics - assignment 4

Johanna Becher, Talha Rehman

## 1 Theory exercises

### 1.1 Task 1

Mutations are changes in a cells genetic information. They can be classified by their formation or by their impact they have.
They can form by insertion or deletion or the change of a single base in the DNA. But it can also happen that a sequence of bases is inserted or deleted.
If the mutation only consists of one base there are diffrent classifications of the mutation depending on the amino acid that is build with the triplet containing the mutation.
A silent mutation occurs when the change of the base doesn't affect the generation of the amino acid at all. A missense mutation leads to a change of the amino acid and a nonsense mutation causes the build in of an early stop codon.

### 1.2 Task 2

Bacteria multiply quite fast by copying it's DNA and dividing into 2 new cells. During the copying mechanism there is a risk that errors occur which would cause a mutation.
Those mutations can have a lethal impact or they can have even no impact at all. But if the mutation causes an adaption of the bacteria to a given condition, due to Darwin's theory of natural selection, it is very propable that this bacteria is going to grow better.
If a bacteria culture is exposed to an antibiotic it could happen, that somme bacteria mutate and their descendants become resistant to this antibiotic.

### 1.3 Task 3

When it comes to modelling of 3 dimensional protein structure we want to be able to compare different models and how "good" they are. Therefore we compare our models to experimantally resolved protein structures by using a measure for structure similarity.

One of those measures is called RMSD, which calculates the root mean square deviation of the C$\alpha$-atoms of n superimposed amino acids.

$$RMSD(v,w) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(||v_i - w_i||^2)}$$

Where $v = (v_1, .., v_n)$ is the first protein with n amino acids, w the second protein and $vi = (v_{ix}, v_{iy}, v_{iz})$ consists of the coordinates of the i-th C$\alpha$-atom in v (respectively w).
Since RMSD weights all atoms equally it is very sensitive to local structure deviations, but it does not take into account the length of the alignment. This means the shorter the alignment is, the better the RMSD will be.

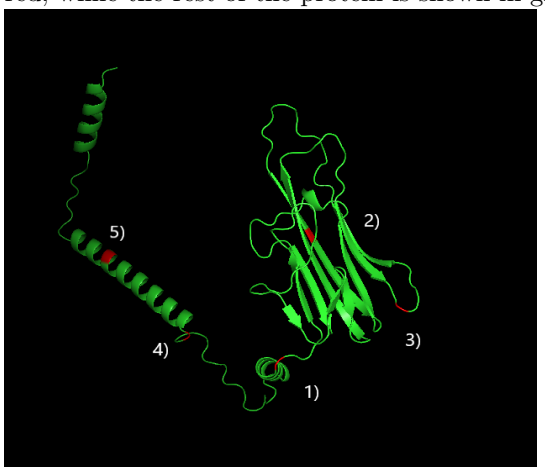
Another measure, called GDT_TS (=global distance test, total score) calculates the largest set of C$\alpha$-atoms of two superimposed protein structures. It calculates a percentage of how many C$\alpha$-atoms have a distance smaller than a predefinded cutoff.
It is more accurate than the RMSD and is used as a major assessment criteria in the CASP.

## 1.4 Task 4 and 5

We were looking at the protein *TNF-alpha* which is secreted by activated macrophages. It causes the necrosis of tumors (what the name already indicates) but it also plays a role in different signaling pathways.
The top 5 missense mutations are the following:

| Variant ID | Source | HGVS Consequence | VEP Annotation | Allele Count | Allele Number | Allele Frequency |
|---|---|---|---|---|---|---|
| 6-31544562-C-T | E G | p.Pro84Leu | missense | 674 | 278064 | 2.42e-3 |
| 6-31545193-T-A | E G | p.Ile194Asn | missense | 94 | 277956 | 3.38e-4 |
| 6-31545093-G-A | E G | p.Val161Ile | missense | 30 | 278418 | 1.08e-4 |
| 6-31543679-G-A | E G | p.Gly54Glu | missense | 17 | 255450 | 6.65e-5 |
| 6-31543619-G-C | G | p.Ser34Thr | missense | 2 | 31390 | 6.37e-5 |

In the following picture you see the positions of those 5 mutations marked in red, while the rest of the protein is shown in green.




The numbers indicate the mutations by their allele frequency,

1) position 84: Pro $\rightarrow$ Leu

2) position 194: Ile $\rightarrow$ Asn

3) position 161: Val $\rightarrow$ Ile

4) position 54: Gly $\rightarrow$ Glu

5) position 34: Ser $\rightarrow$ Thr

Since mutations 2) and 5) are insinde a secondary structure they could break up this structure, what could cause a loss of functionality.
Changing Isoleucine into Aspargine could cause some changes. Both amino acids have a similar size but different side chains.
Replacing Serine with Threonine would only add a $CH_3$ group what wouldn't have a big impact on the $\alpha$-helix.
The other 3 muatations are placed in loops connectiong secondaary structures.
The first mutation looks quite interesting, because it is quite close to an helix. When you replace Proline with Leucine you get an amino acids which prefers helices more than loops. It could cause an elongation of the $\alpha$-helix.
The same could happen at mutation 4) because Glutamic Acid also likes helices.

Those changes of the secondary strcture were not detectable with SwissModel, neither for mutation 4) nor mutation 1).